

Mixing time of generalized Rosenthal walk on $SO(n)$

Yunjiang Jiang

October 26, 2011

Abstract

We prove that a uniformized variant of both the Rosenthal walk [10] and the Kac random walk [6] on $SO(n)$ mixes in $\mathcal{O}(n^3)$ steps in total variation distance. The proof also extends easily to Rosenthal walk with fixed angle $\theta \neq \pi$. To the best of our knowledge, this is the first polynomial time bound for both walks. The techniques employed are mainly from representation theory of $SO(n)$. But a crucial new ingredient is the interpretation of the Fourier coefficients of the character ratio as counting the number of particle cascade paths arising from the classical branching rules.

Consider the random walk on $SO(N)$ defined by $X_0 = \text{id}$, $X_j = X_{j-1}S_j$, where S_j are iid random elements in $SO(N)$ uniformly supported on the conjugacy class of $R(1, 2; \theta)$. In other words, each step we take a random rotation of angle θ along a uniformly chosen two-plane. The goal is to bound the total variation mixing time to its stationary distribution, namely the Haar measure.

For two probability measures on the same space Ω , recall the following consequence of the Cauchy-Schwarz inequality:

$$\begin{aligned} \|\mu - \nu\|_{\text{TV}}^2 &= \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathcal{L}^1(d\nu)}^2 \\ &\leq \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathcal{L}^2(d\nu)}^2. \end{aligned}$$

Here we are assuming the density $\frac{d\mu}{d\nu}$ exists, but even when it doesn't, the first equality still holds if we restrict the integral to where $\frac{d\mu}{d\nu} < \infty$. However, the \mathcal{L}^2 norm might be infinite.

Now we specialize to $\Omega = SO(N)$ and $\nu = U$, the Haar measure. Then by Plancherel's theorem, the right hand side equals (see [5] Chapter 3)

$$\sum_{\rho \in \widehat{SO(N)}} \text{Tr}(\widehat{\mu - U})(\rho)(\widehat{\mu - U})(\rho)^T = \sum_{\rho \neq 1} \text{Tr} \widehat{\mu}(\rho) \widehat{\mu}(\rho)^T.$$

To see the last equality, simply observe that for $\rho = 1$, and any probability measure π , $\widehat{\pi}(\rho) = 1$, hence $\widehat{(\mu - U)}(\rho) = 0$; for any other ρ , $\widehat{U}(\rho) = 0$ since $\rho(g)\widehat{U}(\rho) = \widehat{U}(\rho)$ for any $g \in SO(N)$, by left-invariance of U .

Furthermore, we have the convolution identity $\widehat{\mu * \nu}(\rho) = \widehat{\mu}(\rho)\widehat{\nu}(\rho)$. Now let S_j 's be distributed according to μ . Then since μ is constant on conjugacy classes, by Schur's lemma, we know that $\widehat{\mu}(\rho)$ is always a constant times the identity matrix. Therefore

$$\text{Tr} \widehat{\mu^{*t}}(\rho) \widehat{\mu^{*t}}(\rho)^T = d_\rho^2 (c_\rho/d_\rho)^{2t},$$

where $c_\rho = \text{Tr} \widehat{\mu}(\rho)$.

Now we further restrict to the case of $N = 2n + 1$. Then the irreducible representations of $SO(N)$ are indexed by n -tuples $(0 \leq a_1 \leq a_2 \leq \dots \leq a_n)$ [7]. We will denote d_a and $c_a = c_a(\theta)$ the dimension and the character value evaluated at $R(1, 2; \theta)$ of the irreducible representation labeled by the n -tuple a . Note that Rosenthal [10], [1] used the following slightly different labeling convention $(a_1 + 1/2, a_2 + 3/2, \dots, a_n + n - 1/2)$. He gave the following explicit formulae for d_a and c_a by taking suitable limits in the Weyl character formula:

$$d_a = \frac{2^n}{1!3! \dots (2n-1)!} \prod_{q=1}^n (a_q + q - 1/2) \prod_{1 \leq s < r \leq n} [(a_r + r - 1/2)^2 - (a_s + s - 1/2)^2]$$

$$c_a(\pi) = \frac{(2n-1)!}{(2 \sin(\theta/2))^{2n-1}} \sum_{j=1}^n \frac{\sin(a_j + j - 1/2)\theta}{(a_j + j - 1/2) \prod_{r \neq j} (a_r + r - 1/2)^2 - (a_j + j - 1/2)^2}$$

By restricting ρ_a to the canonical copy of $SO(2)$ in $SO(N)$, it decomposes into a non-negative linear combination of irreducible representations of $SO(2)$, which are of the form $t \mapsto t^k$ for integer k , $|t| = 1$. When $k = 0$, we have the trivial representation. Now since $R(1, 2; \theta)$ is an element of $SO(2) \subset SO(N)$, we can express $\rho_a(R(1, 2; \theta))$ in terms of irreducible representations of $SO(2)$, so for $\alpha_j \geq 0$,

$$\rho_a(R(1, 2; \theta)) = \sum_{j=0}^m \alpha_j \cos(j\theta).$$

Clearly, $c_a(\pi) = \sum_j \alpha_j (-1)^j \geq \alpha_0 - \alpha_1$. The α_j 's admit the following interpretation. By Weyl's character formula, one can express the restriction of ρ_a from $SO(N)$ to $SO(N-1)$ as follows:

$$\rho_a^{(2n+1)}|_{SO(2n)} = \bigoplus_{|b_1| \leq a_1 \leq b_2 \leq \dots \leq b_n \leq a_n} \rho_b^{(2n)}$$

$$\rho_b^{(2n)}|_{SO(2n-1)} = \bigoplus_{|b_1| \leq a_1 \leq b_2 \leq \dots \leq b_n} \rho_a^{(2n-1)}$$

so each irreducible representation of $SO(N-1)$ appears at most once when restricted from $SO(N)$. This yields a combinatorial description of α_j , namely the number of shrinking ensemble paths with the particle non-intersecting condition as given above such that the topmost particle arrives at $\pm j$ in $N-2$ steps (since we are going from $SO(N)$ to $SO(2)$). We have the following two immediate consequences:

Lemma 0.1. $\alpha_1 \geq \alpha_2 \geq \dots$, so the α_j are monotone non-increasing starting at $j = 1$.

Proof. We will let β_j be the number of ensemble paths that leads to the $a_1^{(3)} = j$ at the $SO(3)$ level. Then condition on $a_1^{(3)} = j$, $j \geq 1$, the next step one could get $b_1^{(2)} \in [-j, j]$, each with multiplicity 1. Thus if $1 \leq r < s \in \mathbb{Z}$, the conditional multiplicity of $a^{(2)} = \pm r$ is no less than that of $a^{(2)} = \pm s$. This shows $\alpha_r \geq \alpha_s$. Since there is only one copy of $a^{(2)} = 0$, α_0 is not necessarily $\geq \alpha_1$. \square

Lemma 0.2. *Let β_j be as defined in the previous proof. Then $2\beta_j = \alpha_j - \alpha_{j+1}$ for $j \geq 1$, and $\beta_0 = \alpha_0 - \frac{1}{2}\alpha_1$. Furthermore we have $\beta_j \geq \frac{2j+1}{2j+3}\beta_{j+1}$.*

Proof. The relation between β and α follows from the following two equations:

$$\begin{aligned}\alpha_0 &= \sum_{i=0}^m \beta_i \\ \alpha_j &= 2 \sum_{i=j}^m \beta_i.\end{aligned}$$

These are consequences of the interlacing condition for the branching rule from $SO(3)$ to $SO(2)$. Next we condition on $a_1^{(5)} \leq b_2^{(4)} \leq a_2^{(5)}$. Then $b_1^{(4)}$ can be any integer in the range $[-a_1^{(5)}, a_1^{(5)}]$, each with multiplicity 1. So the number of branching paths down to $SO(3)$ with $a_1^{(3)} = j$ is given by $\beta'_j := [1 + 2(j \wedge a_1^{(5)})]I_{\{j \leq b_2^{(4)}\}}$, where $\beta'_j := \beta_j^{a_1^{(5)}, b_2^{(4)}, a_2^{(5)}}$ is the conditional β so to speak. Then it is clear by case analysis that

$$\beta'_j \geq \frac{1+2j}{3+2j}\beta'_{j+1}.$$

Now summing over all the conditions $a_1^{(5)} \leq b_2^{(4)} \leq a_2^{(5)}$ we obtain the second assertion. \square

Observe now that $\sum_{j \geq 0} \alpha_j = d_a$, since each shrinking ensemble path represents one copy of some irreducible representation of $SO(2)$, which are all one-dimensional. We shall use the notation $\tilde{\alpha}_j = \alpha_j/d_a$, and similarly $\tilde{\beta}_j = \beta_j/d_a$. Note that $\sum_{j \geq 0} \tilde{\beta}_j \neq 1$.

The last lemma implies the following bound:

Lemma 0.3. *If $\tilde{\alpha}_0 > \tilde{\alpha}_1$, then $\tilde{\alpha}_0 \leq \mathcal{O}((\tilde{\alpha}_0 - \tilde{\alpha}_1)^{1/3}) \leq \mathcal{O}(\sum_{j \geq 0} \tilde{\alpha}_j (-1)^j)$.*

Unfortunately, the condition in the above lemma only holds for small representations.

Proof. Since $\alpha_j - \alpha_{j+1} = 2\beta_j \leq 2(1+2j)\beta_0$ by telescoping, we have

$$\alpha_0 - \alpha_{j+1} = \alpha_0 - \alpha_1 + \alpha_1 - \alpha_2 + \dots + \alpha_j - \alpha_{j+1} \leq 3 \sum_{i=0}^j (1+2i)\beta_0,$$

i.e., $\alpha_{j+1} \geq \alpha_0 - (2j+3)j\beta_0$. Now using the fact that $\sum_{j \geq 0} \tilde{\alpha}_j = 1$, we have for all $k \geq 0$,

$$(k+1)\tilde{\alpha}_0 - \sum_{j=1}^k (2j+3)j\tilde{\beta}_0 = (k+1)\tilde{\alpha}_0 - f(k)\tilde{\beta}_0 \leq 1,$$

where $f(k)$ has cubic growth. So

$$\tilde{\alpha}_0 \leq \frac{1}{k+1} + \frac{f(k)}{k+1} \tilde{\beta}_0.$$

Optimizing over k on the right hand side, we obtain

$$\begin{aligned} \tilde{\alpha}_0 &\leq c(\tilde{\beta}_0)^{1/3} \\ &= c(\tilde{\alpha}_0 - \frac{1}{2}\tilde{\alpha}_1)^{1/3} \\ &\leq c(\tilde{\alpha}_0 - \tilde{\alpha}_1)^{1/3} \\ &\leq c(\sum_{j \geq 1} \tilde{\alpha}_j (-1)^j)^{1/3}. \end{aligned}$$

Where in the last inequality we used monotonicity of α_j , $j \geq 1$. □

Next we come back to bounding the total variation distance using the character sum:

$$\|\mu^{*t} - U\|_{\text{TV}} \leq \sum_{a \neq (0^n)} d_a^2 \left(\frac{c_a}{d_a}\right)^{2t},$$

where (0^n) is the label for the trivial representation. We first give an upper bound on d_a . Since we know $d_{(0^n)} = 1$, it suffices to look at the ratio:

$$\frac{d_a}{d_{(0^n)}} = \prod_{q=1}^n \frac{a_q + q - 1/2}{q - 1/2} \prod_{r>s} \frac{(a_r + r - 1/2)^2 - (a_s + s - 1/2)^2}{(r - 1/2)^2 - (s - 1/2)^2}$$

Now suppose $a = (0^{n-k}, a_{n-k+1}, \dots, a_n)$, i.e., the first $n-k$ coordinates are all 0. Then we have the following bound:

$$\begin{aligned} \frac{d_a}{d_{(0^n)}} &= \prod_{j=n-k+1}^n \frac{a_j + j - 1/2}{j - 1/2} \prod_{s=1}^{n-k} \prod_{r=n-k+1}^n \frac{a_r + r - s}{r - s} \frac{a_r + r + s - 1}{r + s - 1} \\ &\leq c \prod_{r=n-k+1}^n \exp(a_r \sum_{s=1}^{n-k} \frac{1}{r-s}) \exp(a_r \sum_{s=1}^{n-k} \frac{1}{r+s}) \\ &\leq c \prod_{r=n-k+1}^n \left(\frac{r-1}{r-n+k}\right)^{a_r} \left(\frac{r+n-k}{r+1}\right)^{a_r} \\ &\leq c \frac{n^{\sum_{r=n-k+1}^n a_r}}{\prod_{j=1}^k j^{a_{n-k+j}}} \end{aligned}$$

1 Upper bound on the mixing time of the uniform Rosenthal walk

Consider the following random walk on $SO(n)$. Each step, an independent uniform angle $\theta \in [0, 2\pi)$ is chosen and then an independent and uniform element in the conjugacy class of

$R(1, 2; \theta)$ is chosen. The walk then multiplies itself by this doubly stochastic element to reach the next step. This is closely related to the Kac random walk on $SO(n)$, where instead of choosing a random element in the conjugacy class, one chooses a random pair of coordinates $1 \leq i < j \leq n$, and multiplies the chain by the random matrix $R(i, j; \theta)$. In the Kac case, no polynomial mixing time upper bound is known [8] (though see [9], [4], and [3] for results in other modes of convergence). Here we present a polynomial mixing time argument for the corresponding Rosenthal walk.

First we need a simple lemma

Lemma 1.1. *Consider a random walk K generated by μ on a compact group G which is constant on conjugacy classes and another one J generated by ν which can be of any type (doesn't even need to be ergodic). Then the \mathcal{L}^2 distance to stationarity satisfies the following censoring inequality:*

$$\|\mu^{t_1} * \nu * \mu^{t_2} * \dots * \nu * \mu^{t_k} - U\|_2 \leq \|K^t - U\|_2.$$

where $\sum_i t_i = t$ and $t_i \geq 0$.

Proof. This simply follows from the Plancherel's identity. It clearly suffices to show that for each irreducible representation ρ of G ,

$$\mathrm{Tr} \mu^{t_1} * \widehat{\nu} * \mu^{t_2}(\rho) \leq \mathrm{Tr} \hat{\mu}^t(\rho).$$

But this is true simply because $\hat{\mu}(\rho) = cI$ for some constant c so can be collected in the front of the final matrix product. \square

Now we will consider two auxilliary random walks to the uniform Rosenthal walk. Let K be the Rosenthal walk with θ uniformly supported on $[\epsilon, 2\pi - \epsilon]$, and J be the unconditional walk. if we condition the Rosenthal walk on θ in the range $[\epsilon, 2\pi - \epsilon]$, then we obtain K above. So the strategy now will be to show that K mixes in $O(n^2 \log n)$ steps and that with high probability in $O(n^2 \log n)$ steps there will be that many steps where K is taken.

The second claim is easy to deal with. Let $\#K$ denote the number of steps in which $\theta \in [\epsilon, 2\pi - \epsilon]$.

$$\mathbb{P}_{ct}[\#K \geq t] = \sum_{j=t}^{ct} \binom{ct}{j} (1-\epsilon)^j \epsilon^{ct-j}.$$

For c sufficiently large, the distribution of $\#K$ converges to a Gaussian centered at $ct(1-2\epsilon)$ with standard deviation about \sqrt{ct} . So if we choose $c > 2$ then certainly $\#K \geq t$ with overwhelming probability.

So now it suffices to bound the \mathcal{L}^2 mixing time of K .

- Regime 1. When $r(\pi) = \sum_{j=0}^{\infty} \alpha_j (-1)^j > 1/6$, we can infer that $\alpha_0 > \alpha_1$, so the full monotonicity of the Fourier coefficients holds. Let

$$T_a(j) := \frac{(2n-1)!}{2^{2n-1}} \frac{(-1)^j}{(a_j + j - 1/2) \prod_{r \neq j} ((a_r + r - 1/2)^2 - (a_j + j - 1/2)^2)}.$$

From the Rosenthal bound on $r(\pi)$,

$$r_a(\pi) \leq \sum_{j=1}^n T_a(j),$$

we see that if $a_n > 12n$, then $r_a(\pi) < 1/6$. This is because the probability measure on $[n]$ defined by

$$\mu(j) := T_{0^n}(j) = \frac{(2n-1)!}{2^{2n-1}} \frac{(-1)^j}{(j-1/2) \prod_{r \neq j} [(r-1/2)^2 - (j-1/2)^2]},$$

is concentrated near $j = 1$ with a Gaussian decay of window size $n^{1/2}$, as easily confirmed by Stirling approximation. Furthermore for any $a \in (\mathbb{N}^n)^\uparrow$, $T_a(j) \leq \mu(j)$ using the monotonicity $a_0 \leq a_1 \leq \dots \leq a_n$. Therefore $\sum_{j > cn^{1/2}} T_a(j) = o(1)$, and we only need to focus on $j < cn^{1/2}$. For the latter we simply compare $T_a(j)$ with $\mu(j)$ and bound the ratio uniformly:

$$\frac{(a_j + j - 1/2)^2 - (a_n + n - 1/2)^2}{(j - 1/2)^2 - (n - 1/2)^2} \geq \frac{a_j + a_n + j + n - 1}{j + n - 1}.$$

So if $a_n > 12n$, this ratio is > 6 . In fact this is true for all $j \in [n]$. Therefore by monotonicity of expectation, we easily get $r_a(\pi) < 1/6$.

Now observe that a_n is precisely the number of nonzero α_j 's in the Fourier expansion of $r(\theta)$, using the branching rule characterization. Therefore by monotonicity of α_j 's,

$$\begin{aligned} \alpha_1 &\geq \frac{\sum_{j=1}^{a_n} \alpha_j}{a_n} \\ &\geq \frac{\alpha_1 + \alpha_3 + \dots}{a_n} \\ &\geq \frac{1}{2} \frac{1 - r(\pi)}{a_n}. \end{aligned}$$

Finally we have for $\cos \theta < 1 - \epsilon$ (which is the case with our assumption that θ is bounded away from $0 \bmod 2\pi$), that

$$r(\theta) < 1 - (1 - \cos \theta) \alpha_1 < 1 - (1 - \cos \theta) \frac{1 - r(\pi)}{2a_n}.$$

Now if we let A be all the irreducible representation indices of $SO(2n+1)$ with the property that $r_a(\pi) > 1/6$, then Rosenthal clearly implies that

$$\sum_{a \in A} d_a r_a(\pi)^{n \log n} = o(1),$$

which in turn gives

$$\sum_{a \in A} d_a r_a(\theta)^{\frac{2a_n}{1 - \cos \theta} n \log n} = o(1).$$

- Regime 2. From Rosenthal's work [10], we know the character ratio at a particular θ is bounded as follows:

$$r(\theta) \leq \frac{(2n-1)!}{(2 \sin \theta/2)^{2n-1}} \sum_{j=1}^n \frac{1}{a_j |\prod_{r \neq j} (a_r + r)^2 - (a_j + j)^2|}.$$

Let us denote $W_s = \prod_{r>s} |\frac{(a_r+r)^2 - (a_s+s)^2}{r^2 - s^2}|$. Using the fact that the character ratio is 1 for the trivial representation, taking the smallest W_0 gives a bound of $r(\theta)$ as follows:

$$r(\theta) \leq \frac{1}{(\sin \theta/2)^{2n-1}} \sum_{j=1}^n \frac{1}{W_0}.$$

We also have the control over the gap between the smallest and largest W_s , $\max_s W_s \leq (\min_s W_s)^{2n}$. This follows from the simple and crude observation that for any fixed j ,

$$\begin{aligned} (a_n + a_j + n + j) &\geq \frac{1}{2}(a_r + a_s + r + s) \\ (a_n + a_j + n + j) &\geq (a_r - a_s + r - s). \end{aligned}$$

Actually the bound above is the best one could hope for, as it is achieved by the following family of representations $(0^{n-1}, k)$, as $k \rightarrow \infty$.

But now we have $r(\theta) \leq \frac{1}{(\sin \theta/2)^{2n-1}} (\min_j W_j)^{-1} n$, whereas the dimension $d \leq (\max_j W_j)^{2n/2}$ by throwing away the corresponding terms for the trivial representation, and estimating the factor $\prod_q (a_q + q - 1/2)$ trivially by the product of the W_j 's. So denoting by W_0 the minimizer of W_j , we need t such that the following expression is small:

$$[W_0^{2n}]^n [\sin^{-(2n-1)}[\theta/2] W_0^{-1} n]^t$$

It is important that our upper bound for $r(\theta)$ is < 1 . to accomplish that we need $W_0 \geq e^{\Omega(n)}$. In that event, we certainly can take $t = \Omega(n^2)$ and get the required bound.

- Regime 3. Now we need to consider the case when our upper bound for $r(\theta) \geq 1$. One possibility is taken care of by Regime 1. So it suffices to consider the situation where $r(\pi) < 1/6$. But that means $\alpha_0 < \frac{7}{12}$ by an easy computation using the fact that $\alpha_0 - \alpha_1 + \alpha_2 \leq r(\pi)$. Integrating over $\theta \in [\epsilon, 2\pi - \epsilon]$, we obtain

$$\frac{1}{2\pi - 2\epsilon} \int_{\epsilon}^{2\pi - \epsilon} r(\theta) d\theta = \alpha_0 - \sum_{j=1}^{\infty} \frac{1}{j(\pi - \epsilon)} \sin(\epsilon j).$$

It is clear that $\frac{\sin \epsilon j}{j(\pi - \epsilon)} \leq \frac{1}{\pi - \epsilon}$.

So for fixed $\epsilon \in (0, \pi - 1)$, we have the uniform bound $\int_{[\epsilon, 2\pi - \epsilon]} r(\theta) d\theta \leq 1 - c_{\epsilon}(1 - \alpha_0)$. So in other words, $r(\theta) < 1 - \delta$ for some $\delta > 0$. On the other hand, the dimension is bounded by

$$d \leq (W_0^{2n})^n \leq \exp(\mathcal{O}(n^3)).$$

So taking $t = \Omega(n^3)$ suffices to kill all the harmonics in this regime.

In summary, the mixing time of the uniform Rosenthal walk is $\mathcal{O}(n^3)$. There is a corresponding lower bound in \mathcal{L}^2 of order n^2 (see Porod and Rosenthal). The above argument clearly shows that the mixing time of the uniform walk is of the same order as for a fixed θ (or uniform θ is a region bounded away from 0). So the conjectured total variation mixing time is still $n \log n$.

2 Acknowledgement

I received generous help from Bob Hough in making the connection of Fourier expansion of the character ratio with restriction of irreducible representations as well as some preliminary bounds for the character ratio. Thanks are due to Daniel Bump [2] for explaining to me how the classical branching rule for $SO(n)$ works. I would also like to thank Persi Diaconis and Arun Ram for help with literature search.

References

- [1] J. Frank Adams. *Lectures on Lie groups*. W. A. Benjamin, Inc., New York-Amsterdam, 1969.
- [2] Daniel Bump. *Lie groups*, volume 225 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2004.
- [3] Eric A. Carlen, Maria C. Carvalho, Jonathan Le Roux, Michael Loss, and Cédric Villani. Entropy and chaos in the Kac model. *Kinet. Relat. Models*, 3(1):85–122, 2010.
- [4] Eric A. Carlen, Jeffrey S. Geronimo, and Michael Loss. Determination of the spectral gap in the Kac model for physical momentum and energy-conserving collisions. *SIAM J. Math. Anal.*, 40(1):327–364, 2008.
- [5] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [6] Persi Diaconis and Laurent Saloff-Coste. Bounds for Kac’s master equation. *Comm. Math. Phys.*, 209(3):729–755, 2000.
- [7] Roe Goodman and Nolan R. Wallach. *Representations and invariants of the classical groups*, volume 68 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1998.
- [8] Yunjiang Jiang. Total variation bounds of kac random walk. *Ann. Appl. Probab.*, 2011.
- [9] Roberto Imbuzeiro Oliveira. On the convergence to equilibrium of Kac’s random walk on matrices. *Ann. Appl. Probab.*, 19(3):1200–1231, 2009.
- [10] Jeffrey S. Rosenthal. Random rotations: characters and random walks on $SO(N)$. *Ann. Probab.*, 22(1):398–423, 1994.